4-30-2019

Analyzing and Estimating Cyberattack Trends by Performing Data Mining on a Cybersecurity Data Set

Chan Young Koh
Bridgewater State University

ANALYZING AND ESTIMATING CYBERATTACK TRENDS BY PERFORMING DATA MINING ON A CYBERSECURITY DATA SET

Chan Young Koh

Submitted in Partial Completion of the
Requirements for Departmental Honors in Computer Science

Bridgewater State University

April 30, 2019

_____
Dr. Enping Li, Thesis Advisor

_____
Dr. Laura Gross, Committee Member

_____
Dr. Haleh Khojasteh, Committee Member

Acknowledgement

I would like to acknowledge and extend my heartfelt gratitude to the following people who have made the completion of this thesis possible:

My mentor, Dr. Enping Li, for her consistent encouragement and thorough guidance throughout the research.

Dr. Laura Gross and Dr. Haleh Khojasteh for gladly reviewing and showing interest to my research.

Dr. Wanchunzi Yu, who helped me to understand many useful fundamental concepts of the language R I used throughout my research.

All Computer Science faculty members and Staff, who have taught me so much that facilitated the process of this research.

To my family and my girlfriend Anna, who motivated me with unconditional love and support.

And most especially to God, who made all things possible.

Table of Contents

ACKNOWLEDGEMENT **2**

ABSTRACT **4**

LIST OF DEFINITIONS **5**

1. INTRODUCTION **7**

2. BACKGROUND **8**

2.1 Cyber Security *8*

2.2 Cyber Crimes *9*

3. RELATED WORKS **10**

4. DATA MINING **11**

4.1 Data Selection *12*

4.2 Data Inspection *13*

4.3 Data Wrangling *14*

4.4 Data Analysis *15*

4.4.1 General Analysis *15*

4.4.2 Selective Analysis *20*

5. COUNTERMEASURES **21**

5.1 Countermeasures to Malware *22*

5.2 Countermeasures to Account Hijacking *23*

5.3 Countermeasures to SQL Injection *25*

5.4 Countermeasures to DDoS *27*

5.5 Countermeasures to Targeted Attacks *27*

6. CONCLUSION **28**

TABLE OF FIGURES **29**

BIBLIOGRAPHY **30**

Abstract

More than five billion personal information has been compromised over the past eight years through data breaches from notable companies, and the damage related to cybercrime is expected to reach six trillion USD annually by the year of 2021 [1]. Interestingly, recent cyberattacks were aimed specifically at credit agencies and companies that hold credit information of their customers and employees. The question is: “Why is it difficult to protect against or evade cyberattacks even for these prestigious companies?”. The purpose of this research is to bring the notion of notorious, rapidly-multiplying cyberthreats. Hence, the research focuses on analyzing cyberattack techniques and finding effectiveness of surveillance methods that companies utilize to protect themselves from cyberattacks. In order to achieve this, we selected cyberattacks information and analyzed the data set through data mining, and the research findings suggest a future trend of cyberattacks efficient countermeasures. From the information gathered through data mining, the research findings suggest a future trend of cyberattacks and efficient countermeasures.

List of Definitions

1) **Account Hijacking**: The process through which any person's or organization's account(s) - be their email, social media or other accounts - is/are stolen or hijacked by a hacker.
2) **Malware**: Malware is short for "malicious software" - computer programs designed to infiltrate and damage computers without the user's consent.
3) **Ransomware**: Ransomware is a malware that restricts users from accessing their computer systems. To grant access to their restricted data, victims must pay money through online payment methods. Some ransomware encrypts files and is referred to as Cryptolocker.
4) **PoS Malware**: Point-of-Sale malware targets consumers' personal and financial data stored on either credit or debit cards. Hacker will hack the Point-of-Sale machine and use RAM scrapers to collect and sell customer information.
5) **DDoS**: A distributed denial-of-service (DDoS) attack is one of the most powerful weapons on the internet. DDoS attacks target websites and online services. DDoS inundates its target's server capacity with unusual, high network traffic, causing malfunction and immobilization.
6) **SQL Injection**: SQL Injection or SQLi is a type of database threats, in which an attacker uses webpages to plant the attack. Hackers would include malicious SQL query statements in the web server. One loophole in the server may send these malicious SQL query statements into the database itself, causing a deletion of entire database or a severe malfunction.

7) **Whitelist**: A whitelist is a cybersecurity list, only giving administrator-approved programs, and IP and email address, system assess. Any information not on the list will be blocked.
8) **Multi-factor Authentication**: An authentication method which grants access to a computer user after the user successfully presents two or more validation components.
9) **Targeted Attack**: A type of cyberthreat in which perpetrators actively pursue and compromise a target entity's infrastructure while maintaining anonymity. These attackers have a certain level of expertise and have sufficient resources to conduct their schemes over a long-term period. They can adapt, adjust, or improve their attacks to counter their victim's defenses.
10) **Bot**: An application that performs an automated task. Though bots can be beneficial in technology, malicious bots cause serious harm to the computer system.
11) **Data Mining**: A technology for knowledge-discovery in databases, which performs the modeling of a large amount of data to discover relationships among data. find particular data patterns and derive data tendency.
12) **Cybersecurity**: A technology used to protect the data, access, and integrity of computing assets against cyberattacks.

1. Introduction

Even at this very moment, a vast number of cyberattacks occur unpredictably. In 2013, Utah's secure government network reported to have more than 20 million cyberattacks each day. Considering an insane amount of cyberattacks happened in just one state out of many, it is certainly not an exaggeration to suggest that cyberattack is one of the most common crimes that exists in the world today. In that year, Target was affected by a data breach through cyberattack, resulting in 41 million consumers data compromised. In 2015 Anthem, a leading insurance giant, lost their accountability when they realized that about 80 million people's records which included Social Security Number (SSN) and Driver License were leaked by hackers. Later [3], Uber, Forever21, and even LinkedIn admitted that they, too, suffered as data breach victims [4]. Even governmental organizations, Republican Party (GOP) and U.S. Department of Defense, could not prevent themselves from data breach.

These cases conveyed that any organization or company is predisposed to any form of cyberattacks. Incidents of cyberattacks caught much attention from the public not only due to the gravity of the circumstances, but also with the unprofessionalism companies portrayed as a reaction to cyberattacks. Rather than reporting the data breach and its damage at its occurrence, companies concealed the information, leaving their entire customer information compromised.

In fact, most cyberattack cases initiated from tiny faults or loopholes in a computer or a network system. Thus, it is important for the society to be aware of cyberthreats that are proven to be ubiquitous [3]. Since cyberthreats are inevitable, it's rather effective to understand their attack vectors and to prepare for countermeasures if an attack occurs.

In this thesis, we observed a three-years-worth (2016-2018) of cyberattack data set using the statistics programming language R to perform data mining. Through the usage of R, we were able to produce useful analysis. Details regarding the data analysis and its procedure are explained throughout the work.

2. Background

This section is a background study of the following areas: Cybersecurity and Cyber Crimes.

2.1 Cybersecurity

Cybersecurity is a practice that ensures the integrity, confidentiality and availability (ICA) of information. The scope of cyber security is widespread, and a good cyber security strategy should take infrastructure, cloud service, application, and IoT (Internet of Things) securities into account [5].

Cybersecurity has never been more streamlined. As attackers evolve day-to-day, it is important to properly define cybersecurity and identify the conditions of an effective cyber security environment. The importance of cybersecurity is portrayed by the steady increase in the budgets that are spent worldwide on cybersecurity. Solely focusing on the spending on cybersecurity in the United States between 2010 and 2018, the total spending underwent a drastic growth from 27.4 billion US dollars to 66 billion US dollars (an increase of 140.88%) as shown in **Figure 1**. Organizations have begun to recognize that because malware is publicly available as a commodity, anyone can become a cyber attacker and companies that provide security solutions are growing, but there are many solutions that do not do much in terms of providing effective solutions to cyberattacks. Likewise, cybersecurity requires focus and dedication [2].

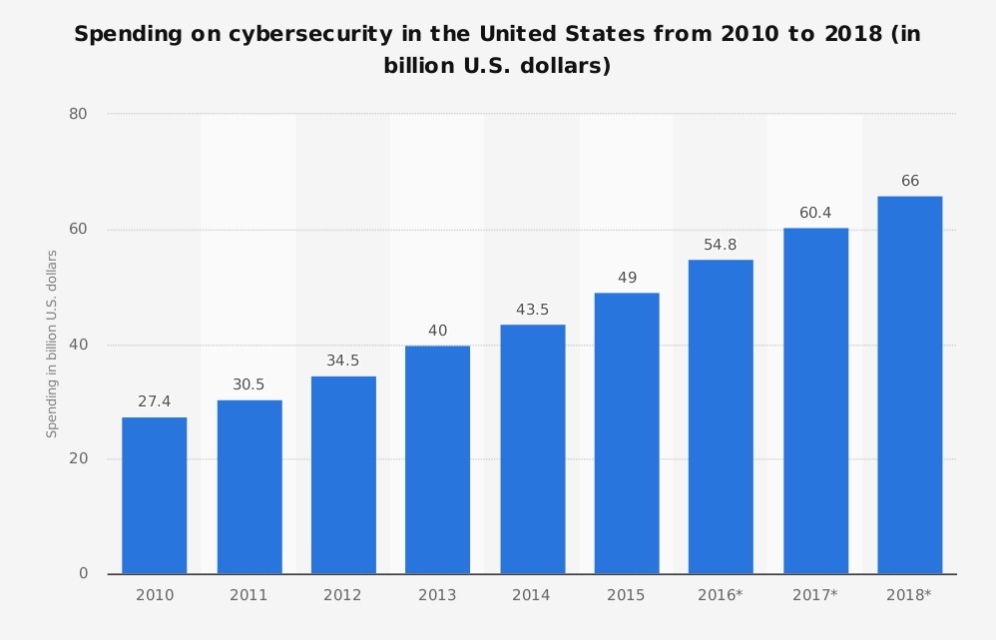

Figure 1. *Spending on cybersecurity in the U.S. (Source: Telecommunications Industry Association, (2015), Statista).*

2.2 Cyber Crimes

Cybercrimes can be recognized as crimes that target a computer system connected to an information network, such as the Internet. Though cybercrimes exist in the domain of cyberspace, the number of cybercrimes increased exponentially due to the development of IT devices and mobile communications and the popularization of Internet use in the past years. Specifically, between 2016 and 2018, cybercrime was statistically denoted as the major motivation of cyberattacks. Cybersecurity statistics have suggested that cybercrime is responsible for 75% of cyberattacks as depicted in **Figure 2**.

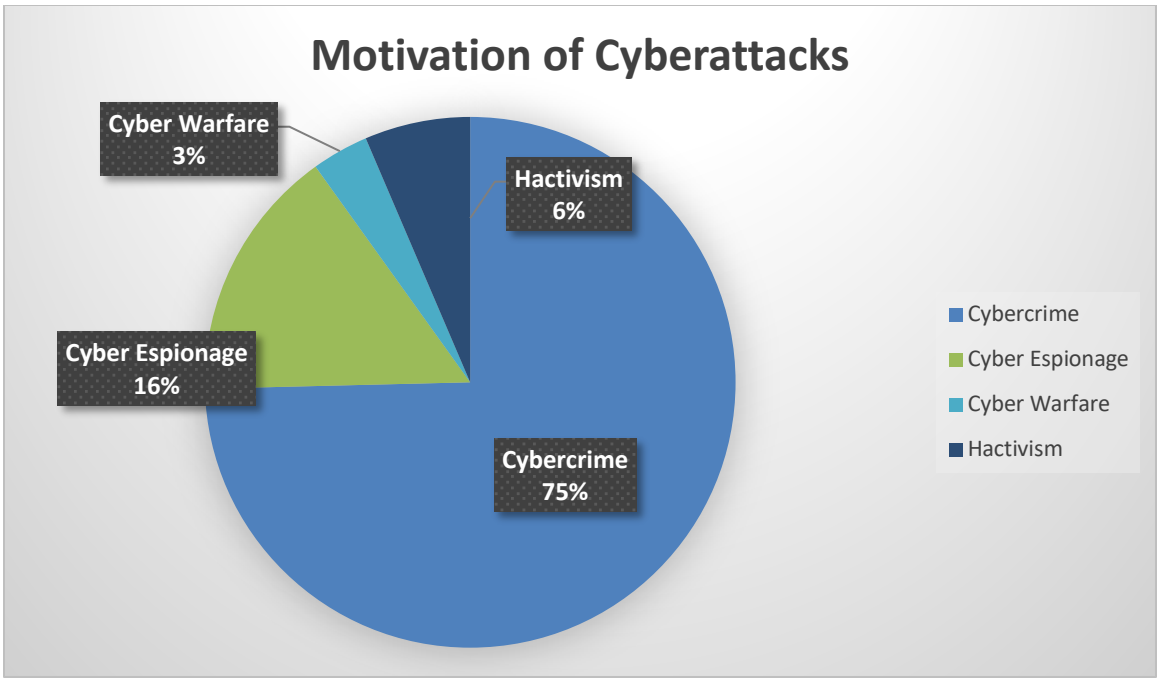

***Figure 2**. Motivation of Cyberattacks (Source: Hackmageddon, (2018), 2016-2018 Master Table).*

The main reason why cyberattacks are popular is because there is no limitation of time or space for the perpetrators as long as they are connected to an information network. A network system is always predisposed to cyberattacks. Some [6] attacks may happen within a short timeframe where as some attacks may take about four months to prepare for a single, deadly attack [7].

3. Related Work

Though the history of data mining is relatively new, there are many research and surveys that denote the effectiveness and efficiency of data mining, especially for cyberattack detection. The author of this thesis acknowledges the following research and surveys that have contributed an inspiration in this current research.

A leading hardware company, IBM, released a paper in 2010 which discusses how data mining can play a useful role in cybersecurity. IBM claims that through data mining, it is plausible to detect insurance fraud from a company's existing data. A company should

take a full advantage of their historical records to minimize irregular records which frequently are deduced as frauds. The paper further focuses on the importance of deep understanding of existing data and creating data models which could be used to detect future insurance fraud.

Anna L. Buzak and Erhan Guven, both members of IEEE[1], argue in their survey that data mining can be used for cyber analytics in support of intrusion detection. Through data mining, one can apply misuse based detection and anomaly based detection, which are considered pivotal in IDS[2]. They claim that data mining facilitates hybrid detection which uses both misuse and anomaly based detection to minimize false alarm generation and maximize the detection rate.

Although the studies above do present an insight of how data mining serves as a key element in cybersecurity, they do not provide a broader spectrum of what type of cyberattacks netizens are expected to encounter and countermeasures for the cyberattacks. However, we highly appreciate the professional findings each research accomplished. The next section presents an overview of data mining on the cybersecurity data sets.

4. Data Mining

As a contribution into the field of computer science, cybersecurity, and statistics, there were multiple constraints to consider before starting the process of data mining.

4.1 Data Selection

A data set must not be biased and must have credibility. The hardest part about data mining for this research was the matter of where to fetch the data from. Technology rapidly

c

[1] Institute of Electrical and Electronics Engineers.
[2] Intrusion Detection System.

develops along with cyberattacks, and it is not appropriate to select outdated, offline data since the focus of this research is on data sets between 2016 and 2018. A website called Hackmageddon caught high interest during the finding of useful data sets since the author of the blog gathered incidents of cyberattacks from different online domains and made a combination of incidents to create a data set [1]. We were able to get in contact with the author of the website, Paolo Passeri, and we requested if he could distribute the raw data files which contain cyberattack incidents between 2012 and 2018. Once he accepted our request, we discovered that the raw data files were larger than the data sets presented on the website. Each entry had a link to check its authenticity, and each link gave more information about each entry as shown in **Figure 3**.

Attack.Class	Country	Link
H	US	https://www.hackread.com/hackers-shut-down-don...
H	SA	https://www.hackread.com/anonymous-takes-down...
H	SA	https://www.hackread.com/ddos-attack-shuts-down...
H	TH	http://www.scmagazine.com/anonymous-attacks-th...
CC	US	http://www.theregister.co.uk/2016/01/05/linode_re...
CC	CN	https://blogs.mcafee.com/mcafee-labs/sms-phishin...
CW	LB	http://www.independent.co.uk/news/world/middle-e...
CC	BR	http://pastebin.com/UUwD4gd4
CC	US	https://www.hackread.com/forbes-website-dropping...
H	UG	https://www.hackread.com/uganda-high-commissio...

***Figure 3**. Each entry has a link attached as a verification method*

4.2 Data Inspection

Once the data sets were selected, we used a statistics program named R to see if every entry in the data set was in a consistent format. During inspection, we recognized that the data from 2016 were separated into months while data sets from 2017 and 2018 were not.

Hence, we started to combine monthly data sets from 2016. However, we encountered a problem while combining the data sets with an error that the column names of the data sets, we were attempting to combine, were not consistent to each other. The inconsistency occurred because the names of the columns were not uniform throughout. Half of the data sets were identified with a column named "X" whereas the remaining data sets used a column named "ID". Since the name "X" does not describe the purpose of the column name, we decided to change the "X" columns to "ID" to make sure each data set has identical columns (**Figure 4**).

```
colnames(dat1)[colnames(dat1)=="X"] <- "ID"
colnames(dat3)[colnames(dat3)=="X"] <- "ID"
colnames(dat5)[colnames(dat5)=="X"] <- "ID"
colnames(dat7)[colnames(dat7)=="X"] <- "ID"
colnames(dat9)[colnames(dat9)=="X"] <- "ID"
colnames(dat11)[colnames(dat11)=="X"] <- "ID"
colnames(dat13)[colnames(dat13)=="X"] <- "ID"
colnames(dat15)[colnames(dat15)=="X"] <- "ID"
colnames(dat17)[colnames(dat17)=="X"] <- "ID"
colnames(dat19)[colnames(dat19)=="X"] <- "ID"
colnames(dat8)[colnames(dat8)=="X"] <- "ID"

identical(names(dat1), names(dat2))
```

***Figure 4**. Changing name of the columns in R.*

Once the names were changed, we ran a function built in R, "identical(data1, data2)", to confirm that every data set has identical columns (**Figure 4**). Finally, we compared the data sets by year to assure that the data sets have identical attributes (columns) as described in **Figure 5**. In the following sub section, we will discuss data wrangling, and why it was significant to achieve our research findings.

```
> ls(MasterTable2018)
 [1] "Attack"       "Attack.Class" "Author"       "Country"      "Date"         "Description"  "ID"           "Link"
 [9] "Target"       "Target.Class"
> ls(MasterTable2016)
 [1] "Attack"       "Attack.Class" "Author"       "Country"      "Date"         "Description"  "ID"           "Link"         "Target"
[10] "Target.Class"
> ls(MasterTable2017)
 [1] "Attack"       "Attack.Class" "Author"       "Country"      "Date"         "Description"  "ID"           "Link"         "Target"
[10] "Target.Class"
```

***Figure 5**. Data sets with identical columns.*

4.3 Data Wrangling

This sub section defines what data wrangling is and the importance of data wrangling to maintain a good data model. Data wrangling is a method of cleaning and joining unorganized and complex data sets for easy access and analysis [8]. We discovered a necessity of data wrangling in the data set since there were many rows that contain either "Null" or "None" under the attribute "Attack". Through the following lines of R code in **Figure 6**, we were able to eradicate cyberattack entries that had no information about the origin of attack.

```r
# Remove rows that conatain "Unknown" or "Null".
MasterTable2016 <- MasterTable2016[MasterTable2016$Attack != "Unknown",]
MasterTable2016 <- MasterTable2016[MasterTable2016$Attack != "Null",]
MasterTable2017 <- MasterTable2017[MasterTable2017$Attack != "Unknown",]
MasterTable2017 <- MasterTable2017[MasterTable2017$Attack != "Null",]
MasterTable2018 <- MasterTable2018[MasterTable2018$Attack != "Unknown",]
MasterTable2018 <- MasterTable2018[MasterTable2018$Attack != "Null",]
```

***Figure 6**. Data wrangling to remove NULL values.*

Once data wrangling was processed, we inspected the data sets again to confirm that no entry under the attribute named "Attack" had either "Unknown" or "Null" (**Figure 7**). This procedure finally allowed us to have an unbiased, combined data set, which ultimately led us to the most important part of this research: the data analysis. Next sub section presents interesting findings through data analysis.

```
MasterTable[, "Attack", drop = FALSE]
```

2	DDoS
3	DDoS
4	DDoS
5	Account Hijacking
6	Account Hijacking
7	Account Hijacking
8	SQLi
9	Malvertising
10	Defacement
11	SQLi
12	Account Hijacking
13	Malware
14	DDoS
15	Defacement

***Figure** 7. Attribute "Attack" no long contains Null or Unknown.*

4.4 Data Analysis

We conducted two type of data analysis with the dataset. We wanted to inspect the data in a broad perspective, which will provide a trend of cyberattacks in general. Then, we will be breaking down the data into parts to determine which of the cyberattacks are prominent within the target domain.

4.4.1 General Analysis

Since the goal of this research is to find if there exists any trend of cyberattacks, we conjectured that using visual models would facilitate the process. Using ggplot2[3], we were able to produce three pie charts that showed the distribution of cyberattacks between 2016 and 2018. **Figure 8** shows ten cyberattacks that had high frequency values. Both DDoS and SQLi take large proportions of the chart because there were extreme cybercrime incidents involving DDoS and SQLi in 2016.

c

[3] A useful R dependency which allows users to create graphic charts from a data set.

Through SQL injection attacks in 2016, more than 500,000 voter records were stolen, more than four million personal records were exposed, and more than two TB [4] of data were compromised. Government organizations and companies including Canonical (distributor of Ubuntu Linux), MySQL, and Symantec that were expected to protect themselves from SQLi were damaged critically. Aside from SQLi attacks, many suffered through distributed denial of service attacks in 2016.

A notable Domain Name System (DNS) provider, Dyn, was under three phases of DDoS attacks on October 21, 2016. Over thirty companies had server disruptions during the attack periods, causing utter inconvenience to customers. Perpetrators used botnet integrated through a vast number of Internet of Things (IoT) devices that had been infected with Mirai[5] malware. The main cause of the incident was from IoTs users leaving their devices with default passwords, providing a security loophole for attackers to take advantage of [6]. At a glance, cyberattacks in 2016 occurred through SQLi, targeted attack, DDos, defacement, and Point-of Sale (PoS) malware/malware. Next, we inspected data set from 2017.

c

[4] A terabyte is equal to 1024 gigabyte which can contain innumerable amount of records (approx. five millions files).

[5] A malware that infects smart devices that run on ARC processors. Devices infected by Mirai malwares turn into a network remotely controlled bots.

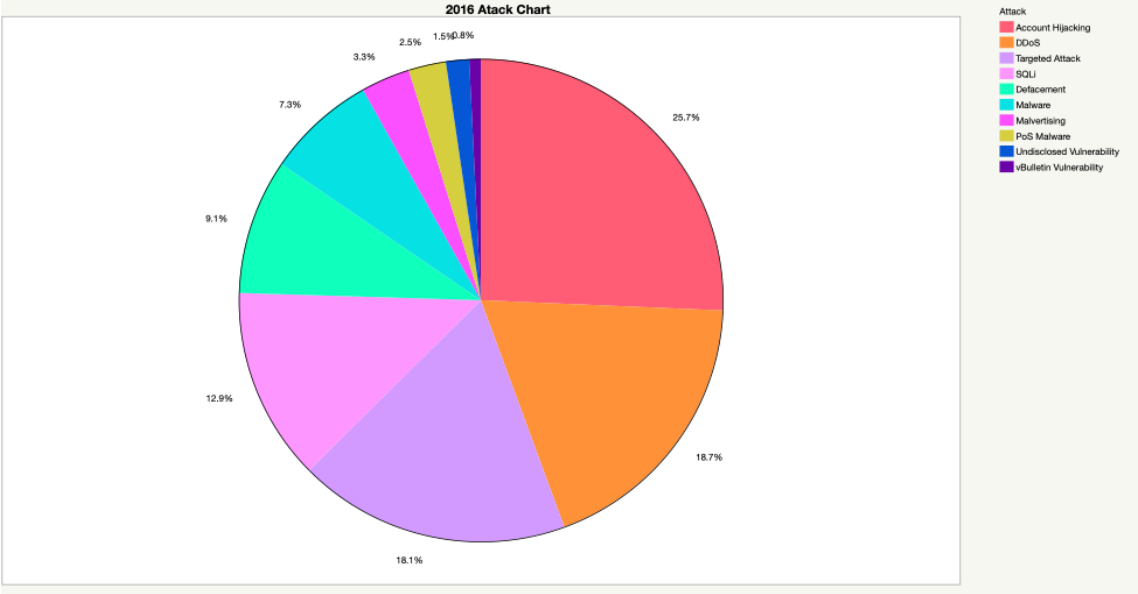

***Figure 8**. 2016 Cyberattack Chart.*

The 2017 data set portrayed a similar trend akin to that of 2016 data set. The most popular six cyberattacks were account hijacking, malware including Point-of-Sale (PoS) malware, targeted attack, DDoS, and Defacement. In **Figure 9**, there is an exponential growth of cyberattacks involving malware. In mid 2017, ransomware cyberattacks went viral internationally. Notorious ransomware such as WannaCry[6], Bad Rabbit[7], and NotPetya[8] damaged over 400,000 machines, and many users and companies lost their files in the computer systems [9]. Aside from ransomware, there was a high number of malwares that disturbed businesses and institutions in 2017. One proposition we had was that the influence of countless data breaches gave leeway to hackers, motivating them to perform more effective cyberattacks through the use of malware.

c

[6] Attack that exploited a flaw in Windows in order to extort money from users.
[7] A ransomware that infected several big Russian and Ukrainian media outlets.
[8] Attack that also exploited a flaw in Windows – known as the EternalBlue.

There were more than two billion items of personal information compromised through revealed data breaches. Prestigious companies such as Equifax, Uber, and Deep Root Analytics were attacked through loopholes in the system [4]. This allowed hackers to use victims' personal information to coin more vicious cyberattacks. Cyberattacks that happened in 2017 have a similar trend to that of cyberattacks happened in 2016. Account hijacking, targeted attack, malware, DDoS, defacement, and PoS malware are predominant. Finally, before determining a future trend of cyberattacks, we observed our data set from 2018.

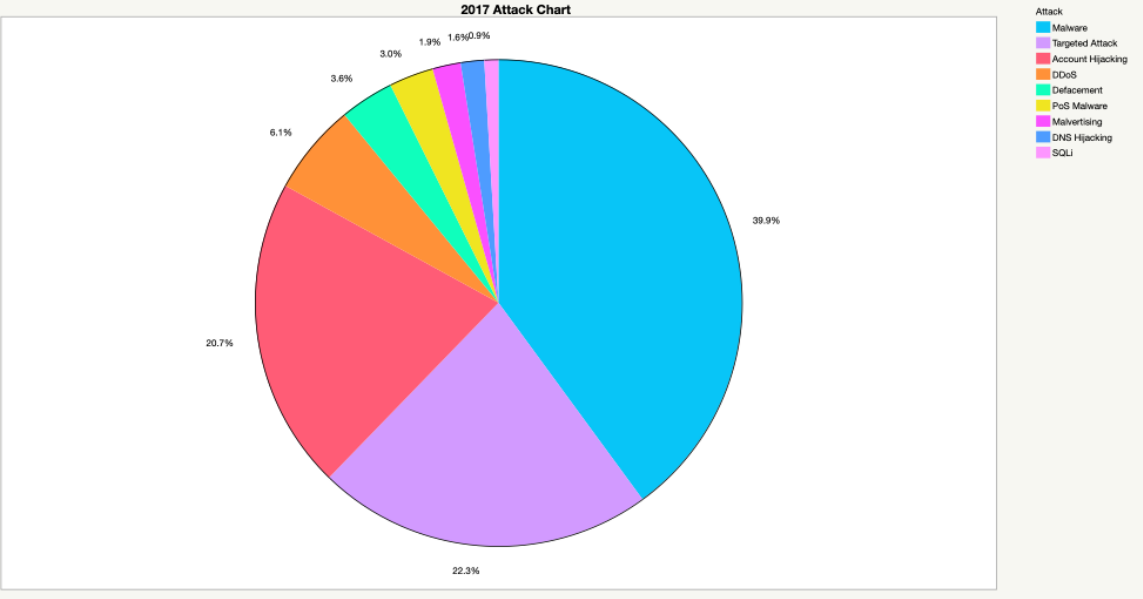

Figure 9. *2017 Cyberattack Chart*

After numerous data breaches in 2017, there were more incidents of data breach that took place in 2018. The Sacramento Bee, a daily newspaper published in Sacramento, California, was attacked by an unknown attacker which resulted in 19.5 million records breaching from their database. Gemini Advisory, the owner of Lord & Taylor; Panera Bread; and Under Armour also suffered with massive data breaches that followed in the year 2018. Facebook was not an exception; In March 2018, about 90 million private records

were breached. The number confirmed that data breaches during 2018 increased by 424% when compared with 2017. In fact, there were countless data breaches in 2018 which created a term “breach fatigue” with the occurrences involving data breaches becoming the “new normal” [10].

These breaches occurred through multiple vulnerabilities in computer and network systems, and detecting these vulnerabilities were difficult since they were tiny faults in the system. In **Figure 10**, account hijacking, targeted attack, and PoS malware or malware continuously take up most of the cyberattacks that occurred in 2018. Overall, it was not easy to consider that these attacks are irrelevant to each other. In fact, we realized that the cause of a certain cyberattack could be another cyberattack, creating a loophole in the computer or network system. For instance, in targeted attacks, criminals use a variation of malware called a spyware to record the victim’s keyboard input in order to gain access to passwords or intellectual property [10].

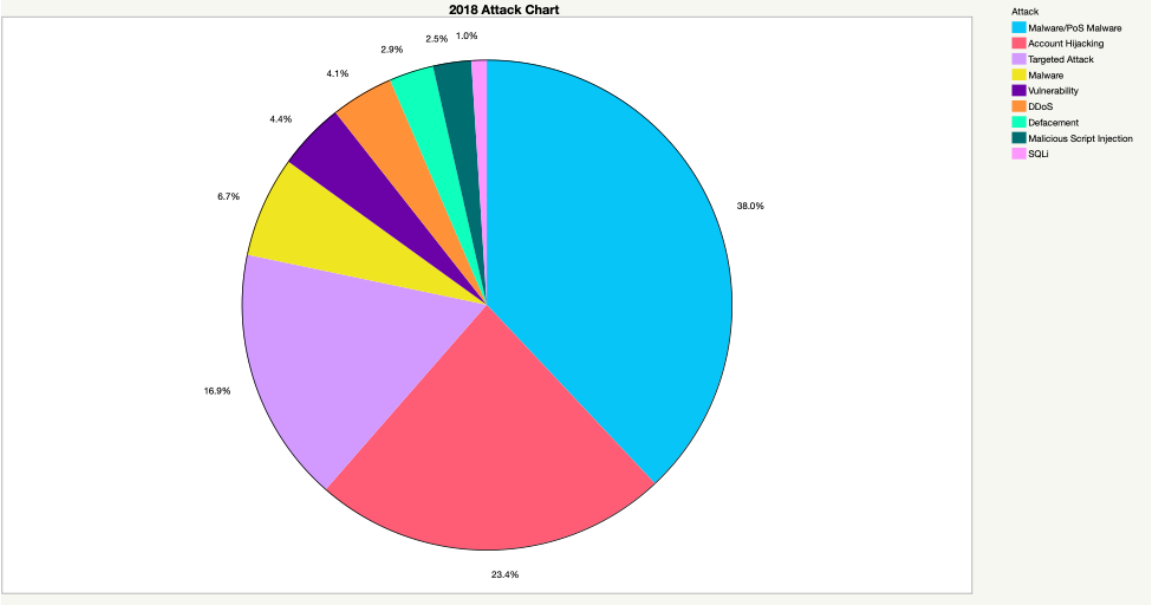

***Figure 10**. 2018 Cyberattack Chart*

4.4.2 Selective Analysis

As stated in section **4.4**, we also observed the data from a specific viewpoint. To provide an unbiased estimation, we decided to order the data set using target classes. We used Pareto plots[9] to determine the difference in cyberattacks within various target classes. **Figure 11** shows how cyberattacks can vary depending on the target they intend to attack.

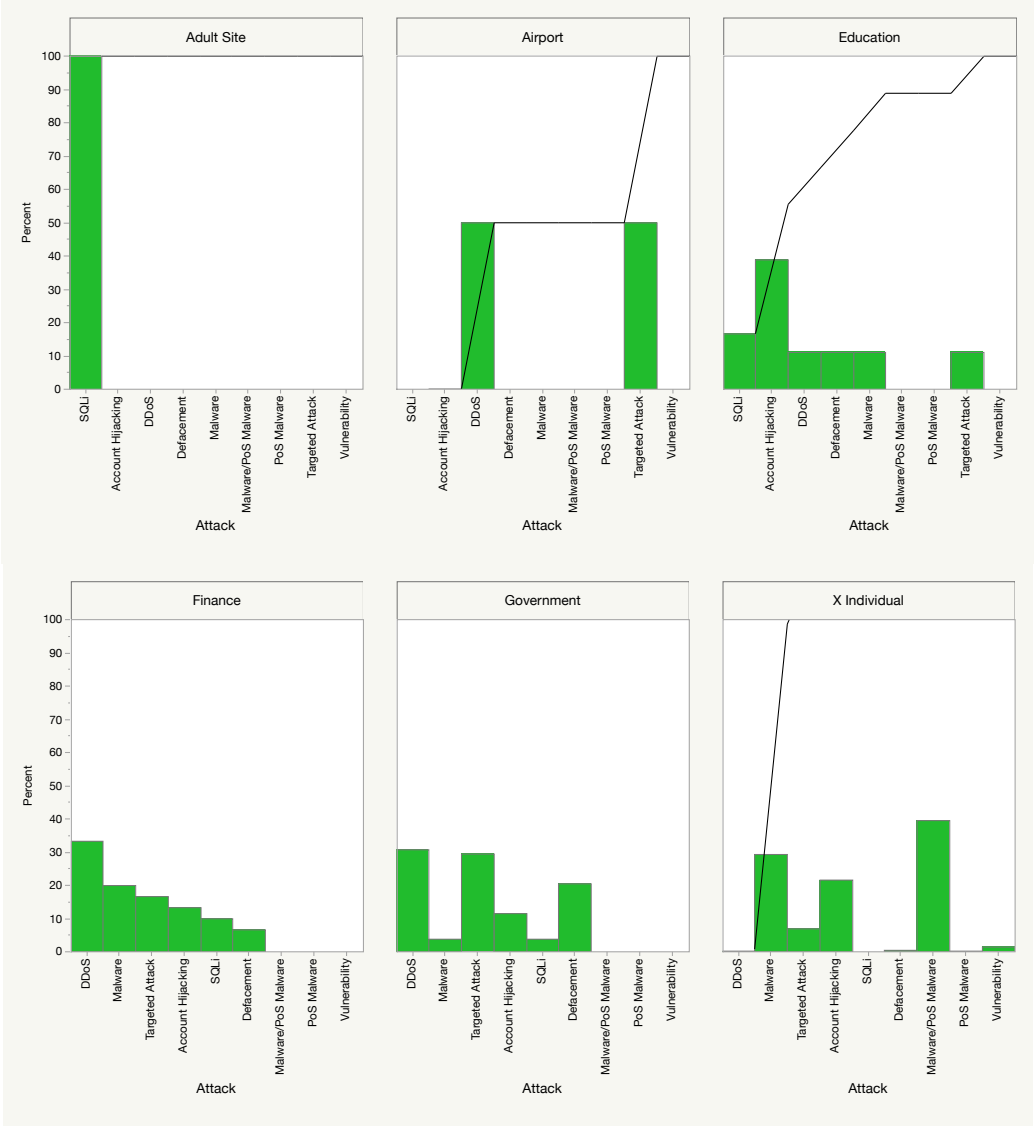

Figure 11*. Pareto plots show that some cyberattacks are different by their target.*

c

[9] A chart that contains both bars and a line graph. Individual values are represented in descending order by bars

It is quite fascinating how cyber criminals know what attack vectors to take depending on their target domains. The information portrayed as Pareto plots also convey that to protect against cyberattacks, people should prepare multifaceted countermeasures in case of attacks. The next section suggests multifarious countermeasures to five, trending cyberattacks.

5. Countermeasures

There is neither a perfect cure for nor prevention of cyberattacks. Rather, it is a fool's errand when one tries to provide a solution to a specific cyberattack because cyber criminals continuously discover new attacks to infiltrate computer and network systems. There are some basic countermeasures that are used commonly by companies and individuals. For instance, cloud

However, concerning incidents like Dyn attack of 2016 which was briefly mentioned **above**, numerous cyberattacks occur from a minor mistake made by a human. Hence, to provide general procedures that would minimize the chance of becoming a victim of cyberattack, this section discusses some basic countermeasures for five most notorious cyberattacks as shown in **Figure 12**: (1) Malware including PoS malware, (2) account hijacking, (3) SQL Injection, (4) DDoS, and (5) targeted attacks [12].

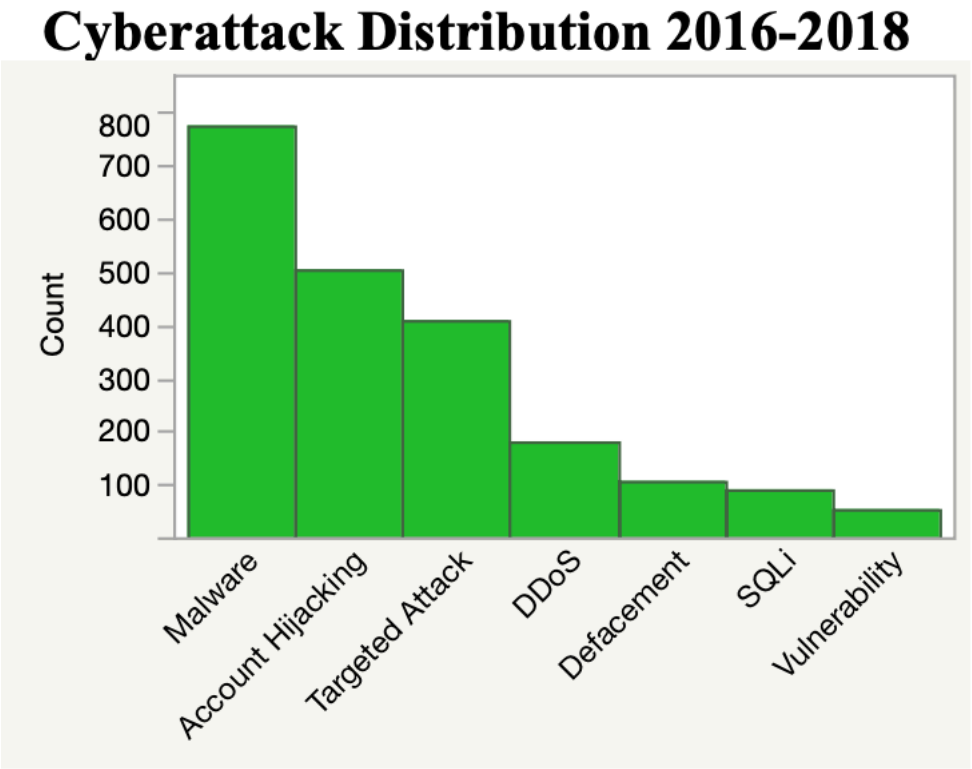

***Figure 12**. Distribution Bar Graph of Cyberattacks.*

5.1 Countermeasures to Malware

There are countless variations of malware today. In fact, AV-TEST Institute[10] claims that they register over 350,000 new malwares per day [11]. This insinuates how maintaining an updated system is crucial. Invigilance of applying important security patches is an automatic invitation for cyber criminals to attack one's computer or network system through malware. Users may also use whitelist to only allow access to certified networks, but whitelisting can bring huge limitation since it blocks every other access [12].

In case of malware attacks, it is wise to encrypt one's file system. Microsoft Windows has the Encryption File System (EFS), which allows users to encrypt their files, making them inaccessible to outside users [13]. Code Signing Certificate, which are used by

c
[10] An independent research institute for IT security from Germany.

software developers, are highly recommended since it allows others to know if the software they are attempting to download is released from a trustworthy provider [14].

The best suggestion to avoid PoS malware is for credit and debit card users is to avoid using their cards in a suspicious area. It is never recommended to carelessly use an unverified automated teller machine (ATM). PoS systems on these ATMs can already be infected with PoS malware. Making a transaction with infected ATMs allows cyber criminals to collect one's credit information from the PoS terminal [1]. For business owners who use PoS systems, whitelisting may be effective. Whitelisting only allows connection with authorized domains in their PoS terminals, which minimizes the chances of malware infection in the system [17].

5.2 Countermeasures to Account Hijacking

The number of cyberattacks caused by account hijacking continues to increase as the world becomes a society of social media. Cyber criminals are fully aware that by hacking into a victim's account, they can scrape many items of personal information which are mostly used to commit another cybercrime. A password may seem to be an ultimate solution to such an issue; however, a more effective countermeasure to project one's online account may be to set up multi-factor authentication [15].

Publicly recognized as two-factor authentication, this method of account protection allows users to set up another layer of security filter with information only the user knows. SMS verification does seem popular nowadays, but it is extremely vulnerable to SIM cloning attacks; therefore, two-factor authentication is recommended above other methods.

Similar to the countermeasure provided for malware, maintaining one's antivirus and network security up-to-date is pivotal. Spyware may be watching of one's hardware system,

trying to scrape any personal information and login credentials. As hardware system protection is of equal concern, it is highly recommended for one to change one's home or business router setting to use WPA2[11] with AES (Advanced Encryption Standard) as an encryption setting. This will make it difficult for cyber criminals difficult to bypass the network security to deploy malware that will be used to hijack accounts. **Figure 13** shows how to set up a network setting to WPA2 with AES on Windows and MacOS.

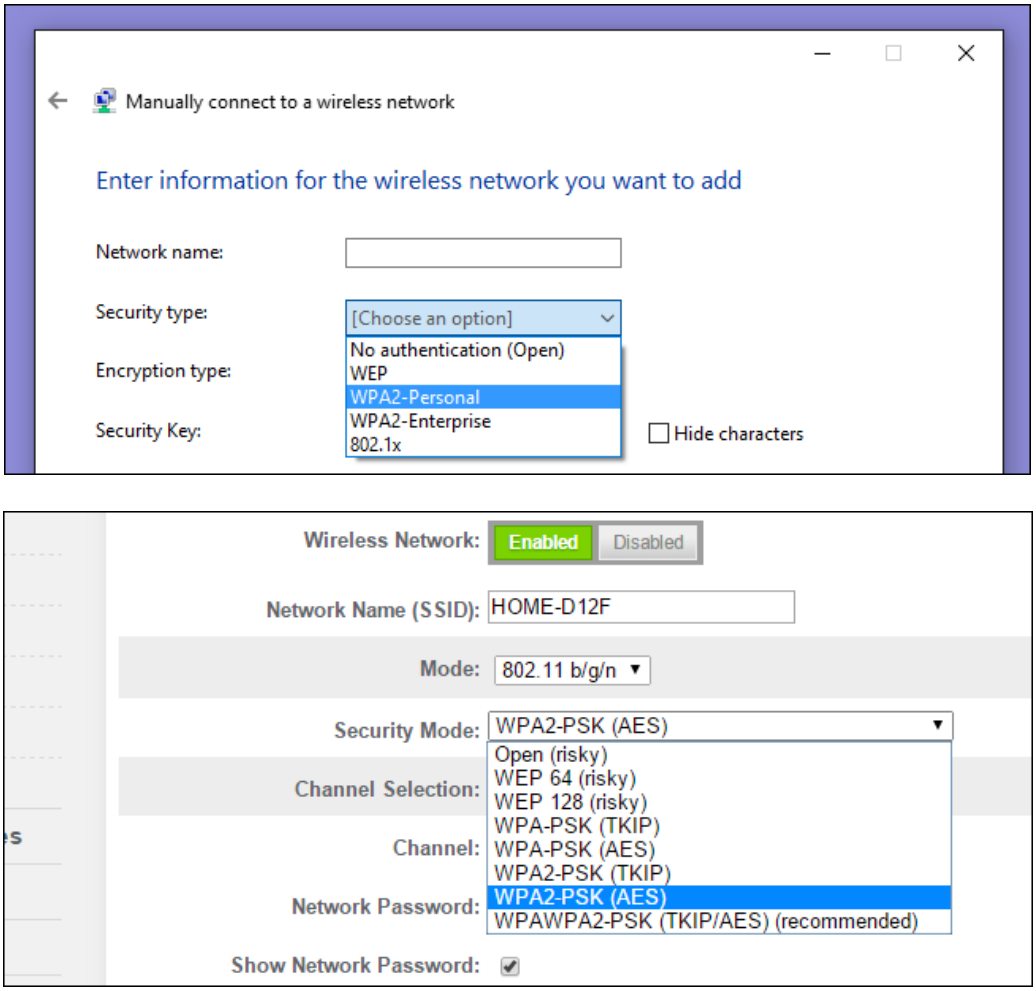

Figure 13. *Setting network security type to WPA2 with AES.*

Many have encountered the sentence, "create a strong password". It is a very common sentence people see online when they try to create a new account. Sometimes, it

c

[11] Wi-Fi Protected Access II is a type of encryption used to secure Wi-Fi networks.

may be difficult to create an effective password that could protect oneself from brute-force automated attacks. **Figure 14** is a simple Python code that checks the length of the password. Once the program confirms that the user has an appropriate (1) length of the password, it will check (2) digits, (3) uppercases, (4) lowercases, and (5) special cases in the password. Note that there is no need of programming to create a strong password, and as long as a person meets these five basic conditions when creating a password, hackers will not be able to simply hack one's account, unless one's system has already been infected with a spyware.

```
def password_check(password):

    # calculating the length
    length_error = len(password) < 8

    # searching for digits
    digit_error = re.search(r"\d", password) is None

    # searching for uppercase
    uppercase_error = re.search(r"[A-Z]", password) is None

    # searching for lowercase
    lowercase_error = re.search(r"[a-z]", password) is None

    # searching for symbols
    symbol_error = re.search(r"\W", password) is None

    # overall result
    password_ok = not ( length_error or digit_error or uppercase_error or lowercase_error or symbol_error )
    return {
        'password_ok' : password_ok,
        'length_error' : length_error,
        'digit_error' : digit_error,
        'uppercase_error' : uppercase_error,
        'lowercase_error' : lowercase_error,
        'symbol_error' : symbol_error,
    }
```

***Figure 14**. Python code to check strong password.*

5.3 Countermeasures to SQL Injection

The best way to avoid SQLi attacks is to utilize prepared statements[12]. Prepared statements protect one's database by only letting authorized personnel know what values are expected in the variable parameter of each SQL statements. Attackers cannot simply coin a malicious SQL statement that would cause a fatal result because if the statement

c
[12] SQL programming method that adds confidentiality within SQL commands.

does not match a specific format, it will simply be ignored and not be queried. Below in **Figure 15**, an example of a prepared statement in provided. We used our data file to query an SQL statement using a prepared statement in R.

```
connection <- dbConnect(RSQLite::SQLite(), ":memory:")

dbWriteTable(connection, "MasterTable", mastertable)

# Using the same query for different values
attack_results <- dbSendQuery(connection, "SELECT * FROM mastertable WHERE [Attack] == ?")
dbBind(attack_results, list("Malware"))
dbFetch(attack_results)
dbBind(attack_results, list("Account Hijacking"))
dbFetch(attack_results)
dbClearResult(attack_results)
```

***Figure 15**. Prepared statement in R.*

Notice that there is a line where "Attack" is equal to "?". This allows the database system to only recognize a specific parameter to be queried, which is only known to authorized personnel. Without prepared statement, attackers can inject queries that could either delete the entire table or steal confidential information from the database. For example, if a parameter is set to an empty string as shown in **Figure 16**, attackers could write, "Account Hijacking; DROP TABLE [some_table_name];". This will delete the entire table without authorization.

```
connection <- dbConnect(RSQLite::SQLite(), ":memory:")

dbWriteTable(connection, "MasterTable", mastertable)

# Using the same query for different values
attack_results <- dbSendQuery(connection, "SELECT * FROM mastertable WHERE [Attack] == ''")
dbBind(attack_results, list("Malware"))
dbFetch(attack_results)
dbBind(attack_results, list("Account Hijacking"))
dbFetch(attack_results)
dbClearResult(attack_results)
```

***Figure 16**. Without prepared statement.*

Aside from prepared statement, monitoring in general is also important. There are various SQL monitoring software that alert issues or changes in the databases before they

are applied, which allows users to check changes before they are made. Monitoring can be still done without a software; however, since human eyes can make mistakes, it is highly recommended to use a software to detect errors that are hard to spot [18].

5.4 Countermeasures to Distributed Denial of Service (DDoS)

Most DDoS starts from bots attacking a computer system. An infected computer will serve as the botmaster[13], and conduct a DDoS attack on other computers. DDoS, just as other cyberattacks, is inevitable, and it is important for anyone to develop a response plan. This task requires a heavy amount of research in DDoS to analyze its attack vectors[14], involving teamwork and effort. Companies should expect that DDoS attacks can occur at any time, and quickly respond to it in case it happens.

To effectively protect one's computer system from DDoS attack, one should definitely set up firewalls, Virtual Private Networks (VPNs), anti-spam, and other layers of protections that could possibly either block or detect attempts of DDoS attacks promptly [16]. There is not one absolute method that would protect one's computer system from DDoS attacks. Instead, it is more appropriate to always be ready for incidents involving DDoS. A well-programmed automated network filtering system could help companies to protect themselves from DDoS attacks.

5.5 Countermeasures to Targeted Attacks

Targeted attacks could simply be regarded as any cyberattacks aimed at a specific organization or an individual. Targeted attacks can use various cyberattacks to infiltrate a computer system. Thus, this complex attack is difficult to avoid for companies that have great security infrastructure. Targeted attacks are mostly specific, meaning, cybercriminals

c

[13] The lead attacking computer in DDoS.

[14] There are three main types of DDoS attacks: Volumetric, Application-Layer, and Protocol attacks.

who commit targeted attacks are well-prepared. This may not be pleasant information to the victims because these perpetrators mostly likely have already analyzed the weakness in their targets' infrastructure. Since there is not one definite solution to targeted attacks, the best reaction to these attacks is to increase the awareness of how targeted attacks can be brutal to others in the related job field [17].

Especially for a company, it is crucial for the managers to educate their workers in preparation of targeted attacks. Once company personnel are aware of the matter, they should deploy an advanced security infrastructure that could effectively protect the company. Network security analytics suggest making segment access within a company's network. By creating segment access, even when hackers successfully infiltrate a part of a company's network infrastructure, they won't have access to other segments [18].

6. Conclusion

To conclude, five cyberattacks, account hijacking, DDoS, SQLi, Malware/PoS malware, and Targeted Attacks, are highly expected to continuously rise in numbers in the future. Specifically, the number of malware will exponentially rise as statistics show that incidents involving malware have increased by 866.7% since 2016.

In regard to countermeasures to these cyberattacks, from the findings, we were able to deduce that the methods to prevent these cyberattacks are similar. Generally, to mitigate each method of cyberattack, update regularly and construct strong security infrastructures. Granted, it is an arduous task for informed netizens to fight in an endless tug-of-war with the cybercriminals. In fact, we can maintain that as technology is gradually becoming more personal than ever, cyberattacks are becoming more personal, targeting sensitive, private information.

Table of Figures

Figure 1. *Spending on cybersecurity in the U.S. (Source: Telecommunications Industry Association, (2015), Statista).* 9

Figure 2. *Motivation of Cyberattacks (Source: Hackmageddon, (2018), 2016-2018 Master Table).* 10

Figure 3. *Each entry has a link attached as a verification method* 12

Figure 4. *Changing name of the columns in R.* 13

Figure 5. *Data sets with identical columns.* 14

Figure 6. *Data wrangling to remove NULL values.* 14

Figure 7. *Attribute "Attack" no long contains Null or Unknown.* 15

Figure 8. *2016 Cyberattack Chart.* 17

Figure 9. *2017 Cyberattack Chart* 18

Figure 10. *2018 Cyberattack Chart* 19

Figure 11. *Pareto plots show that some cyberattacks are different by their target.* 20

Figure 12. *Distribution Bar Graph of Cyberattacks.* 22

Figure 13. *Setting network security type to WPA2 with AES.* 24

Figure 14. *Python code to check strong password* 25

Figure 15. *Prepared statement in R.* 26

Figure 16. *Without prepared statement.* 26

Bibliography

[1] P. Passeri, "Cyber Attacks Statistics," 2018. [Online]. Available: https://www.hackmageddon.com/category/security/cyber-attacks-statistics/.

[2] S. Thielman, "Yahoo hack: 1bn accounts compromised by biggest data breach in history," The Guardian, 15 Dec 2016. [Online]. Available: https://www.theguardian.com/technology/2016/dec/14/yahoo-hack-security-of-one-billion-accounts-breached. [Accessed 2018].

[3] T. Armerding, "The 17 Biggest Data Breaches of the 21st Century," 26 January 2018. [Online]. Available: www.csoonline.com/article/2130877/data-breach/the-biggest-data-breaches-of-the-21st-century.html .

[4] ForcePoint, "What is Cybersecurity?," 2019. [Online]. Available: https://www.forcepoint.com/cyber-edu/cybersecurity.

[5] FireEye, "What is Cyber Security," [Online]. Available: https://www.fireeye.com/current-threats/what-is-cyber-security.html.

[6] S. Hilton, "Dyn Analysis Summary Of Friday October 21 Attack," 26 October 2016. [Online]. Available: https://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/.

[7] Microsoft, "Advanced-Threat-Analytics," 2018. [Online]. Available: https://www.microsoft.com/en-us/enterprise-mobility-security/advanced-threat-analytics.

[8] Altair, "What is Data Wranling?," 2018 . [Online]. Available: https://www.datawatch.com/what-is-data-wrangling/.

[9] ArtsSEC, "Ransomware—Bad Rabbit, Wannacry, and Petya," 17 November 2017. [Online]. Available: https://medium.com/@ArtsSEC/ransomware-bad-rabbit-wannacry-and-petya-26f1725e7778.

[10] Alert Logic, "The 10 Biggest Data Breaches of 2018... So Far," Alert Logic, 16 July 2018. [Online]. Available: https://blog.alertlogic.com/10-biggest-data-breaches-2018-so-far/.

[11] R. A. Grimes, "8 types of malware and how to recognize them," 24 July 2018. [Online]. Available: https://www.csoonline.com/article/2615925/security-your-quick-guide-to-malware-types.html.

[12] K. Kochetkova, "How to not break the Internet," Kaspersky lab , 26 October 2016. [Online]. Available: https://www.kaspersky.com/blog/attack-on-dyn-explained/13325/.

[13] AV-TEST, "Malware," 2019. [Online]. Available: https://www.av-test.org/en/statistics/malware/.

[14] E. Mack, "What Is Whitelisting and How Should You Implement It?," 11 June 2018. [Online]. Available: https://www.springboard.com/blog/what-is-whitelisting/.

[15] Microsoft, "File Encrpytion," 30 5 2018. [Online]. Available: https://docs.microsoft.com/en-us/windows/desktop/fileio/file-encryption.

[16] digicert, "What is Code Signing," 2018. [Online]. Available: https://www.digicert.com/code-signing/.

[17] A. Kumar, "Using whitelisting technology to defend against POS malware," TechTarget, April 2014. [Online]. Available: https://searchsecurity.techtarget.com/answer/Using-whitelisting-technology-to-defend-against-POS-malware.

[18] B. Cha, "What to Do If You'e Been Hacked," 11 September 2014. [Online]. Available: https://www.vox.com/2014/9/11/11630774/what-to-do-if-youve-been-hacked-and-how-to-prevent-it.

[19] T. Keary, "10 Best SQL Server Monitoring Tools for 2019," comparitech, 1 March 2019. [Online]. Available: https://www.comparitech.com/net-admin/sql-server-monitoring-tools/.

[20] B. Dobran, "Seven Tactices To Prevent DDoS Attacks & Keep Your Website Safe," 10 September 2018. [Online]. Available: https://phoenixnap.com/blog/prevent-ddos-attacks.

[21] M. Cobb, "Targeted attack protection: Step-by-step preparation and mitigation," March 2013. [Online]. Available: https://searchsecurity.techtarget.com/tip/Targeted-attack-protection-Step-by-step-preparation-and-mitigation.

[22] S. Deschatres, "The Growth of Targeted Attacks and How to Protect Your Company," 15 April 2014. [Online]. Available: https://www.symantec.com/connect/blogs/growth-targeted-attacks-and-how-protect-your-company.

[23] SAS, "Machin Learning - What it is and why it matters," 2019. [Online]. Available: https://www.sas.com/en_us/insights/analytics/machine-learning.html.